7Б

УДК 519.854:004.8

М.Ф. Бондаренко, В.В. Семенец, Н.В. Белоус, И.В. Куцевич, И.А. Белоус
Харьковский национальный университет радиоэлектроники, г. Харьков, Украина
belous@kture.kharkov.ua

Оценивание тестовых заданий разных типов и определение их уровня сложности

В статье описывается технология оценивания тестовых заданий, а также распределение заданий разных типов по уровням сложности на основе результатов предварительного тестирования. При распределении тестовых заданий по уровням сложности предлагается использовать интегрированную функциональную модель, позволяющую применять разные способы распределения тестовых заданий, включенных в тест, в зависимости от их типов.

Введение

Контроль знаний является важной частью процесса обучения и позволяет получить всестороннюю оценку уровня знаний обучаемых. Одной из хорошо зарекомендовавших себя повсеместно форм контроля знаний является тестирование. Тестирование, несмотря на множество определенных недостатков, является в принципе единственным по-настоящему технологическим средством для измерения уровня знаний и незаменимо как инструмент, позволяющий реализовать эффективный педагогический контроль и должным образом организовать управление учебным процессом.

Одним из сложных и противоречивых вопросов при проведении тестирования является проблема оценивания знаний. Самым распространенным способом решения данной проблемы является использование дихотомической системы оценивания тестовых заданий, в которой за каждое задание можно получить 0 или 1 балл. Данная система удобна при оценивании заданий с выбором одного правильного ответа из многих, т.е. заданий закрытого типа. В то же время существует определенное многообразие типов тестовых заданий: закрытые (многоальтернативные и одноальтернативные), открытые, на установление соответствия между элементами, на установление правильной последовательности, ситуационные тестовые задания [1]. Для оценивания заданий разных типов применение дихотомической системы недостаточно, т.к. обучаемый может дать неполный или частично правильный ответ, что в предложенной системе недостаточно точно оценивается как неправильный ответ. Для эффективного решения данной проблемы необходимо использовать политомическую систему оценивания, в которой допускается несколько категорий ответа на задание, каждая из которых оценивается по-разному. Например, за полностью верный ответ назначается 2 балла, за частично верный – 1 балл и за неверный – 0 баллов. Недостатком этой системы является сложность вычисления общего результата на основе баллов, полученных за задания. Кроме того, в этом случае не учитываются неправильно выбранные варианты ответа. Простое суммирование баллов не соответствует истинному уровню знаний обучаемых. Для решения данной проблемы авторами предлагается введение непрерывной системы оценивания знаний на интервале от 0 до 1 и специализированные технологии определения оценок за выполнение каждого из типов тестовых заданий [2].

Предварительные исследования в области построения системы контроля знаний показали необходимость разделения заданий на уровни сложности. Отсутствие разделения заданий на уровни сложности приводит к недостаточной объективности оценивания знаний и часто не коррелируется с истинным уровнем знаний обучаемых [3]. Так, если сильному студенту попадаются только сложные задания, а слабому – только легкие, то в результате оценивания у обоих студентов будет одинаковый уровень знаний, что не соответствует действительности. И наоборот, если студентам с одинаковым уровнем знаний попадутся задания разного уровня сложности, то проверка знаний выявит у них разный уровень подготовленности, что не является объективным. Распределение заданий по уровням сложности преподавателем вносит субъективизм в процесс оценивания знаний обучаемых по причине того, что не всякое легкое задание для преподавателя является столь же легким и для студентов. Таким образом, разработка технологии распределения тестовых заданий по уровням сложности является актуальной.

Целью работы является разработка технологии проведения тестирования и распределения тестовых заданий по уровням сложности для тестов, оцениваемых по непрерывной системе оценивания знаний, частным случаем которой является дихотомическая система. Для этого авторами предложена методика оценивания тестовых заданий разных форм. Для распределения тестовых заданий по уровням сложности авторами разработана функциональная интегрированная модель современной теории тестов, которая по результатам проведения предварительного контроля знаний определяет способ вычисления уровня сложности тестовых заданий и, после приведения результирующих уровней сложности к единой шкале, выдает устойчивые значения уровней сложности заданий, включенных в тест.

1. Технология оценивания тестовых заданий разных типов

Для создания теста необходимо включать в него задания разных типов. Авторами предлагается при построении теста использовать следующие типы тестовых заданий: закрытого типа (многоальтернативные и одноальтернативные), на установление соответствия, на установление последовательности, открытого типа (введение термина, заполнение таблиц, введение арифметического выражения), ситуационных (авторами предлагается здесь ввести понятие многошагового теста). Для объективной оценки тестовых заданий разных типов предлагается использовать для каждого из них свой специализированный подход для расчета оценки. Введем для определения оценки ответов на задания разных типов коэффициент оценивания r_i.

1.1. Одноальтернативные тестовые задания

Для оценивания одноальтернативного тестового задания достаточно применения дихотомической системы оценивания, где $r_i = 1$ соответствует правильному ответу, $r_i = 0$ – неправильному.

1.2. Многоальтернативные тестовые задания

При оценивании многоальтернативных заданий дихотомической системы недостаточно, потому что обучаемый может дать как неполный ответ, так и один из выбранных вариантов ответа будет неточен. В данном случае необходимо учитывать не

только правильность ответа на задание в целом, но и количество допущенных ошибок. В данном случае коэффициент r_i предлагается рассчитывать по формуле (1).

$$r_i = \frac{Q_{2_s}}{(Q_{1_i} + Q_{3_i})}, \quad (1)$$

где Q_1 – множество всех правильных вариантов ответа в задании,
Q_2 – количество правильных вариантов ответа, выбранных обучаемым,
Q_3 – количество неправильных вариантов ответа обучаемого.

1.3. Задания на установление соответствия

При ответе на задание на установление соответствия каждую пару ответов можно рассматривать как отдельный вариант ответа и при определении результата следует учитывать, сколько пар было выбрано верно. Коэффициент r_i для этого типа заданий рассчитывается по формуле (2).

$$r_i = \frac{Q_{2_i}}{Q_{1_i}}, \quad (2)$$

где Q_1 – количество пар для сопоставления,
Q_2 – количество верно составленных пар.

1.4. Задания на установление последовательности

При оценивании заданий на установление правильной последовательности возможен только один заведомо правильных ответ. Для оценивания данного типа тестовых заданий достаточно использовать дихотомическую шкалу оценивания, коэффициент r_i принимает значение 0 или 1.

1.5. Открытые тестовые задания

При оценивании заданий на введение термина и на введение арифметического выражения достаточно использовать дихотомическую шкалу оценивания, коэффициент r_i принимает значение 0 или 1. При выполнении заданий на заполнение таблиц каждая ячейка таблицы является отдельным вариантом ответа. Если одна из ячеек заполнена неправильно, такой ответ нельзя засчитывать как полностью неправильный (введение одного неверного значения в ячейку может быть механической ошибкой, и поэтому оно должно не полностью обнулить результат выполнения работы, а лишь снизить результат выполнения задания). Для определения коэффициента оценивания заданий на заполнение таблиц рекомендуется использовать показательную функциональную зависимость:

$$r_i = 2^{\frac{Q_{2_i}}{Q_{1_i}}} - 1, \quad (3)$$

где Q_1 – количество ячеек, которые предлагается заполнить обучаемому,
Q_2 – количество ячеек, которое обучаемый заполнил правильно.

1.6. Многошаговые тестовые задания

Многошаговые тестовые задания состоят из набора заданий (набора шагов), решаемых последовательно, когда переход к следующему шагу задания осуществляется только после правильного ответа на предыдущий шаг. Это дает возможность обучае-

мому анализировать не только задание в целом, но и разбираться в каждой составляющей задания. Благодаря этому обучаемый сразу может увидеть, где им допущена ошибка и в дальнейших шагах получить правильные исходные данные, т.е. ошибки в заданиях не будут накапливаться. Многошаговое задание считается пройденным, если на каждом его шаге получен правильный ответ. Для объективного оценивания ответа и глубины знаний обучаемого используется счетчик допускаемых ошибок, количество которых учитывается при выставлении оценки. Кроме того, необходимо решить проблему оценивания таких вопросов. Рекомендуемая формула для вычисления коэффициента правильности для многошагового тестового задания имеет вид :

$$r_i = \frac{\sum_{i=1}^{n} \frac{1}{m_i + 1}}{n}, \qquad (4)$$

где i – номер шага,

m_i – количество ошибок, допущенных на i-м шаге,

n – количество шагов.

Формула (4) справедлива для многошаговых тестовых заданий, в которых на каждом шаге используются одноальтернативные задания или задания на установление правильной последовательности. В случае использования на каком-либо из шагов тестового задания на соответствие или многоальтернативного тестового задания целесообразно использовать формулы (5) и (6) соответственно.

$$r_i = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{Q_{1_{ij}}}{Q_{2_{ij}}}\right), \qquad (5)$$

где j – номер попытки прохождения шага, если на нем была допущена ошибка,

$Q_{2_{ij}}$ – количество пар для составления на i-м шаге при j-й попытке;

$Q_{1_{ij}}$ – количество верно составленных пар на i-м шаге при j-й попытке.

$$r_i = \frac{1}{n}\sum_{i=1}^{n}\frac{m_i \cdot Q_{2_i}}{\sum_{j=1}^{m_i}(Q_{1_{ij}} + Q_{3_{ij}})}, \qquad (6)$$

где $Q_{1_{ij}}$ – количество правильных вариантов ответов на i-м на шаге при j попытке;

Q_{2_i} – количество правильных ответов, выбранных тестируемым на i-м шаге;

$Q_{3_{ij}}$ – количество неправильных ответов, выбранных тестируемым на i-м шаге при j-й попытке.

2. Технология определения результата выполнения теста

В каждой стране для оценивания уровня знаний обучаемых применяется своя, а иногда и несколько шкал оценивания знаний. Так, в Украине применяются 4 различных шкалы оценивания знаний: 100-балльная (рейтинговая), ECTS, 12-балльная и национальная (4-балльная) шкала. Для получения результата в заданной шкале оценивания знаний введем специальный параметр B – балльность системы в виде максимального значения единицы измерения оценки в рассматриваемой шкале. Рассмотрим алгоритм перевода результата тестирования в произвольную шкалу оценивания знаний. Для корректного перевода результата в любую систему оценивания знаний необходимо всем значимым результатам оценивания в порядке возрастания поочередно присвоить коэффициенты балльности, начиная с 1. Максимальное значение коэффициента балльности

и будет соответствовать значению параметра *B*. Для некоторых систем оценивания знаний целесообразно применение интервальной системы перевода результата, в которой в зависимости от процента выполнения работы выставляется оценка.

При оценивании теста, содержащего несколько уровней сложности, каждому из заданий присваивается уровень сложности z_i ($z_i = \overline{1, Z}$). Итоговая оценка выполнения всего теста *R*, состоящего из набора тестовых заданий, содержащего Z уровней сложности, определяется по формуле:

$$R = \frac{B}{\dot{Z}} \times \sum_{i=1}^{N} r_i \times z_i , \tag{7}$$

где *B* – балльность системы,

N – количество тестовых заданий, включенных в тест,

z_i – уровень сложности i-го задания,

$\dot{Z}$ – суммарная сложность теста.

Параметр $\dot{Z}$ является накопительным и определяется по формуле:

$$\dot{Z} = \sum_{i=1}^{N} z_i . \tag{8}$$

3. Технология распределения тестовых заданий разных типов по уровням сложности

Для распределения тестовых заданий по уровням сложности авторами предлагается модификация современной теории тестирования IRT. Модификация IRT заключается в следующем:

– в классическом случае IRT предлагается для вычисления уровня знаний обучаемых. Авторами предлагается применение IRT для определения уровней сложности тестовых заданий;

– модели IRT рассматриваются для случая применения только дихотомической системы оценивания тестовых заданий. Авторами предлагается использование данных моделей для непрерывной системы оценивания знаний, частным случаем которой является дихотомическая;

– предлагается рассматривать одно-, двух- и трехпараметрические модели не в отдельности, а в совокупности, в зависимости от типа тестовых заданий.

Задача распределения тестовых заданий по уровням сложности сводится к определению сложности тестовых заданий с использованием модифицированной современной теории тестирования IRT, исходя из экспериментальных данных предварительного тестирования.

Недостатком современной теории тестирования IRT является наличие трех моделей, каждая из которых в отдельности применима для тестовых заданий определенного типа. При попытке применения модели к разным типам тестовых заданий уменьшается точность вычисления параметров.

Для определения уровня сложности тестовых заданий авторами предлагается интегрированная функциональная модель (рис. 1), рассчитанная на оценивание тестовых заданий по непрерывной шкале оценивания знаний, включающей тестовые задания разных типов:

$$в_j = f(P_{1j}(И_i, res_{ij}), P_{2j}(И_i, a_j, res_{ij}), P_{3j}(И_i, a_j, c_j, res_{ij})), \tag{9}$$

где β_j – параметр, определяющий сложность j-го тестового задания,

$P_{1j}(И_i,res_{ij}),P_{2j}(И_i,a_j,res_{ij}),P_{3j}(И_i,a_j,c_j,res_{ij})$ – модифицированные одно-, двух- и трехпараметрические модели, построенные по непрерывной системе,

Θ_i – параметр, определяющий уровень знаний i-го обучаемого,

res_{ij} – переменная, которая соответствует результату выполнения тестового задания и принимающая значения на интервале от 0 до 1, что соответствует непрерывной шкале оценивания знаний,

a_j – параметр характеристики дифференцирующей способности задания,

c_j – параметр, характеризующий возможность правильного ответа на j-е задание в том случае, если этот ответ угадан.

Достоинством и новизной интегрированной функциональной модели является ее возможность одновременно анализировать тестовые задания всех рассматриваемых выше типов. Технология использования интегрированной функциональной модели заключается в следующем. Тестовое задание T_i, которое подается на вход модели, сначала поступает на блок анализатора типа заданий. Затем, в зависимости от выделенного анализатором типа, задание подается на соответствующий блок определителя результата выполнения тестового задания, где рассчитываются оценки за ответы res_j, а для тестовых заданий закрытого типа дополнительно вычисляется вероятность угадывания правильного ответа c_j.

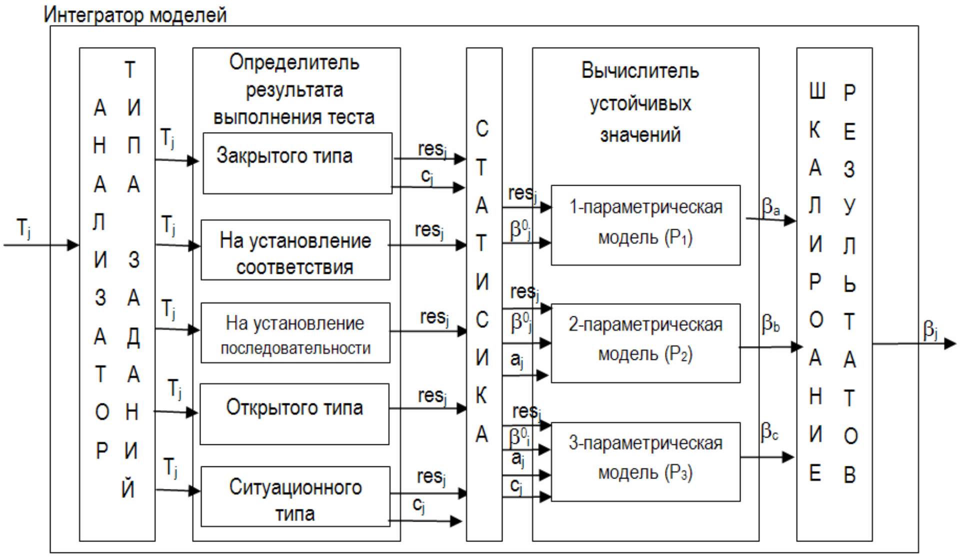


Рисунок 1 – Интегрированная функциональная модель распределения тестовых заданий по уровням сложности

В блоке «Статистика» проводится сбор статистических данных по результатам тестирования знаний обучаемых. После проведения предварительного тестирования группы обучаемых в блоке вычисляется начальный уровень их знаний $И_i^0$, начальный уровень сложности тестовых заданий β_j^0 и рассчитывается дифференцирующая способность тестовых заданий [4]:

$$И_i^0 = \ln\left(\frac{p_i}{q_i}\right),\quad в_j^0 = \ln\left(\frac{p_j}{q_j}\right),\quad a_j = \frac{(r_b)_j}{\sqrt{1-(r_b)_j^2}}, \tag{10}$$

где p_i – доля правильных ответов, полученных от i-го студента,

q_i – доля неправильных ответов, полученных от i-го студента,

p_j – доля правильных ответов, полученных за выполнение j-го задания,

q_j – доля неправильных ответов, полученных за выполнение j-го задания,
r_b – бисериальный коэффициент корреляции.

Далее данные передаются в блок «Вычислитель устойчивых значений», где производится распределение заданий по уровням сложности на основе латентного анализа. После поступления данных в блок «Вычислитель устойчивых значений» производится распределение ответов на тестовые задания и начальных значений параметров для уровней сложности по трем компонентам (**P_1, P_2, P_3**) в зависимости от типа задания.

Первый компонент **P_1** «1-параметрическая модель» применяется для проверки признака гомогенности уровня сложности тестовых заданий относительно гомогенной группы испытуемых. Группа испытуемых является гомогенной, если большинство значений Θ расположено на небольшом интервале оси латентной переменной β. Гомогенный тест представляет собой систему заданий возрастающей трудности, специфической формы и определенного содержания, создаваемую с целью применения объективного, качественного и эффективного метода оценки структуры и измерения уровня подготовленности обучаемых по одной учебной дисциплине. В случае гетерогенной по знаниям выборки испытуемых значения параметра трудности должны охватывать больший интервал оси Θ, а характеристические кривые заданий могут быть расположены довольно далеко друг от друга. Проверка этого условия и реализована в компоненте **P_1** «1-параметрическая модель».

Для повышения точности измерения уровня сложности тестовых заданий вводится параметр характеристики дифференцирующей способности задания a_j, который связан с крутизной кривой задания в точке ее перегиба. При значениях a_j, близких к нулю, тестовые задания утрачивают функцию разделения обучаемых по уровням сложности, что делает их бесполезными при дифференциации по уровню сложности. Число заданий в тесте должно сокращаться в первую очередь за счет удаления таких заданий, что приводит к повышению надежности и валидности теста. Таким образом, для определения уровня сложности β_j тестовых заданий разных типов, за исключением закрытого, служит компонент **P_2** «2-параметрическая модель».

Для тестов с заданиями закрытого типа отмечается существенное отклонение эмпирических данных от теоретической кривой, предсказывающей вероятность правильного выполнения задания при различных значениях переменной Θ. Для решения этой проблемы вводится параметр c_j, который характеризует возможность правильного ответа на задание j в случае, если ответ угадан. Оценки res_j за задания закрытого типа передаются в третий компонент **P_3** «3-параметрическая модель» блока «Вычислитель устойчивых значений».

Для определения устойчивых значений уровней сложности тестовых заданий применяется метод наибольшего правдоподобия Фишера [5], адаптированный к оцениванию знаний обучаемых по непрерывной системе:

$$L_i\{\overline{x_i}|И_i\} = \prod_{j=1}^{n} P_{ij}^{res_{ij}} Q_{ij}^{1-res_{ij}}, \qquad (11)$$

где L_i – вероятностная модель выполнения тестовых заданий для i-го обучаемого;

$\overline{x_i} = \{x_{i1}, x_{i2}, ..., x_{in}\}$ – вектор, характеризующий результат выполнения i-м обучаемым n заданий теста,

P_{ij} – вероятность правильного выполнения i-м испытуемым j-го задания теста,
Q_{ij} – вероятность неправильного выполнения i-м испытуемым j-го задания, $Q_{ij} = 1 - P_{ij}$.

Таким образом, выходными параметрами интегрированной функциональной модели является набор параметров β_j, которые соответствуют устойчивым оценкам уров-

ня сложности тестовых заданий и могут использоваться в тестах, включающих задания нескольких уровней сложности как параметр сложности заданий.

Выводы

Предлагаемая технология реализована программно. Данная технология программно реализована и содержит 2 основных подсистемы:

– модуль проведения контроля знаний, который поддерживает Международный стандарт обмена тестовой информацией [6], сертифицирован [7] и проходит внедрение в высших учебных заведениях Украины;

– модуль распределения тестовых заданий по уровням сложности, который может применяться как при работе с авторским программным комплексом проведения контроля знаний, так и с данными, полученными из внешних систем автоматизированного тестирования. Это делает разработанную систему универсальной.

Соответствующая программная система прошла апробацию на достоверной выборке (более 100 обучаемых).

Внедрение предлагаемой системы в высших учебных заведениях позволит проводить объективное тестирование знаний обучаемых. Применима как в учебных заведениях любого уровня аккредитации, так и в организациях и учреждениях, где проводится профессиональный отбор с помощью тестирования, а также на курсах повышения квалификации, где имеются компьютерные классы.

Литература

1. Комплекс нормативних документів для розробки складових системи стандартів вищої освіти. – Київ, 1998.
2. Belous N. Lifelong education conception using computer testing / N. Belous, I. Voytovich // Материалы VIII Международной конференции Украинской ассоциации дистанционного образования «Образование и виртуальность». – 2004. – С. 307-313.
3. Аванесов В.С. Трудность теста и тестовых заданий / В.С. Аванесов // Управление школой. – 1999. – № 40.
4. Baker F.B. The Basics of Item Response Theory. / Baker F.B. – [2 ed.]– Hieneman, Portsmouth, New Hempshire, 2001. – 174 p.
5. Guilford J.P. Fundamental Statistics in Psychology and Education / Guilford J.P. – N-Y. : McGraw – H.U., 1956.
6. IMS Question & Test Interoperability Specification. [Electronic resource] / IMS Global Learning Consortium: IMS Question & Test Interoperability Specification. – Режим доступа : http://www.imsglobal.org/question/ index.cfm – 15.09.2008 г. – The title from the screen.
7. Белоус Н.В. Программный комплекс для проведения компьютерного и интерактивного обучения и тестирования знаний «КОДЭКС УМА» / Н.В. Белоус, И.В. Войтович // Свидетельство про регистрацию авторского права № 14030 / Государственный департамент интеллектуальной собственности, 02.09.2005.

М.Ф. Бондаренко, В.В. Семенець, Н.В. Білоус, І.В. Куцевич, І.А. Білоус

Оцінювання тестових завдань різних типів та визначення їх рівня складності

У статті розглядається технологія оцінювання тестових завдань, а також розподілення завдань різних типів за рівнями складності на основі результатів попереднього тестування. При розподіленні тестових завдань за рівнями складності запропоновано використовувати інтегровану функціональну модель, яка дозволяє використовувати різноманітні способи розподілення тестових завдань, які було включено до тесту, залежно від їх типів.

M.F. Bondarenko, V.V. Semenets, N.V. Bilous, I.V. Kutsevich, I.A. Bilous

Test Tasks Estimation of Different Types and Definition of their Level of Complexity

The paper is devoted to the test items marking technology and distribution of different types items on complexity levels on the basis of preliminary testing results. At distribution of test items on complexity levels it is offered to use the integrated functional model which allows to apply different ways of the test items distribution included in the test.

Статья поступила в редакцию 06.07.2009.